# BIG DATA AND THE STOCK MARKET: DISTILLING DATA TO IMPROVE STOCK MARKET RETURNS

William Shannon, Jennifer Moranchel and Dr. Xiaoyue Luo
Department of Mathematics, Linfield College, McMinnville, OR 97128

## INTRODUCTION

In our modern competitive market, businesses are seeking efficient and innovative platforms to remain profitable and prepared, especially in the uncertain world of the financial stock market. One possible avenue for improving stock market returns that companies can turn to is harnessing a substantial volume of information, known as big data. However, because of the nature of big data, distilling and analyzing the vast amount of information can require complex analytical methods. Using a keyword selection process based on word frequency, we were able to filter out the data amongst the noise and derive a sector specific keyword list. This list used in combination with a previously created trading method along with the implementation of a thresholding technique, allowed us to develop a more specific trading strategy focused on different market sectors. Our results show that the use of thresholding techniques in addition to the *Google Trends* strategy may improve returns in the stock market.

## *GOOGLE TRENDS* STRATEGY

- If searches go up, take short position. If searches go down, take long position.
  Let current search volume be n(t).
- Step 1: Find the average search volume for the previous $\Delta t$ days.
- Step 2: Determine if n(t) is larger or smaller than this average.
- Step 3: Based on n(t), perform the corresponding trading strategy (taking the short or long position).
- If $\Delta n(\Delta t, t) > 0$, our returns change by $\ln(p(t)) - \ln(p(t+1))$.
- If $\Delta n(\Delta t, t) < 0$, our returns change by $\ln(p(t+1)) - \ln(p(t))$.
- Position is dependant on search volume by using average search volume over previous $\Delta t$ days.
- It is possible to normalize this value over multiple $\Delta t$ values to get a more representative result.

## FIGURES



Figure 1: "Google Trends" strategy for the word "disease" using data from financial market sector of *Google Trends* applied to the S&P 500.
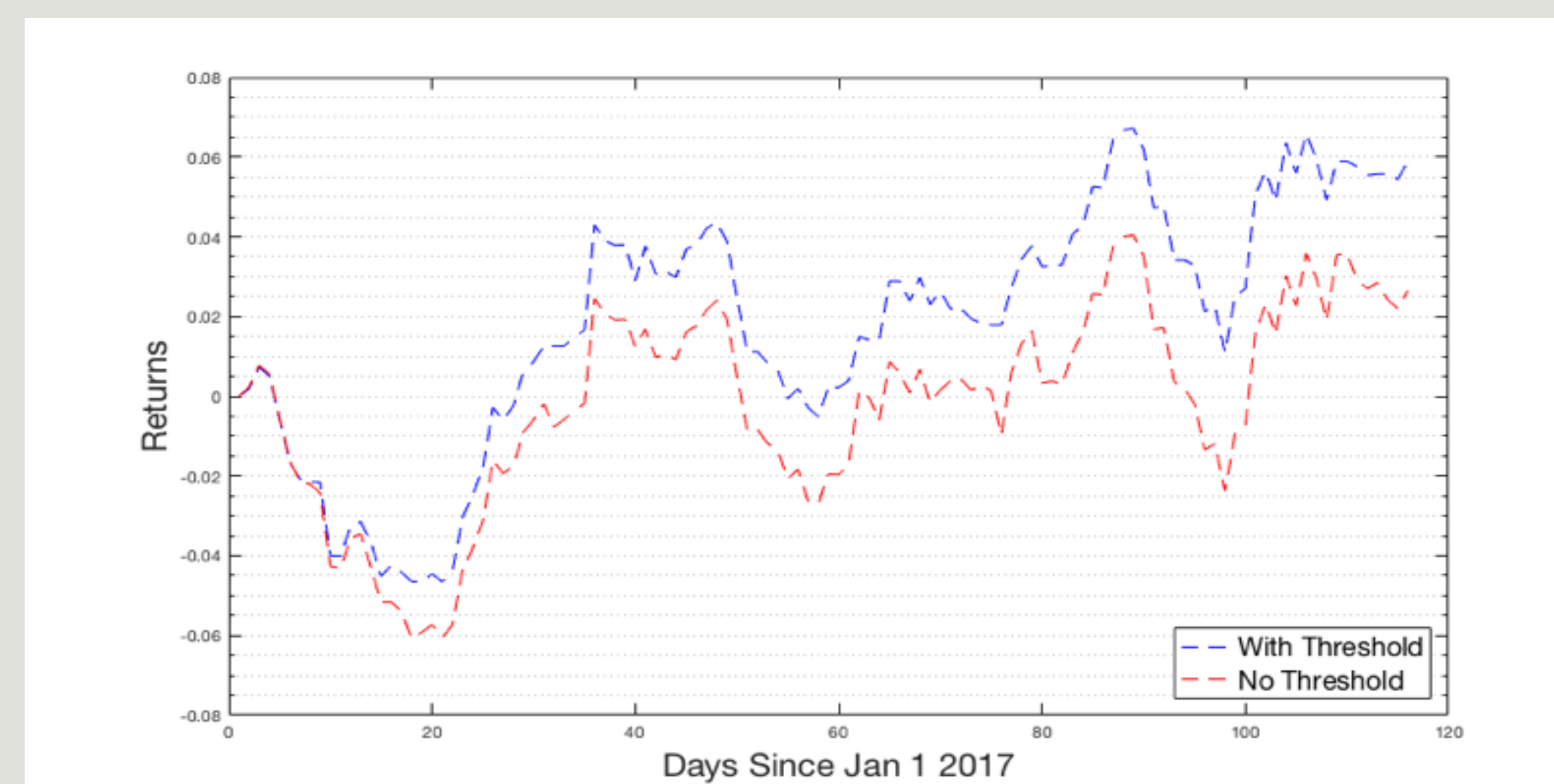


Figure 2. Word "production" with no threshold and with a threshold value of 0.03 applied to energy sector of the S&P 500 using overall *Google Trends* Data.
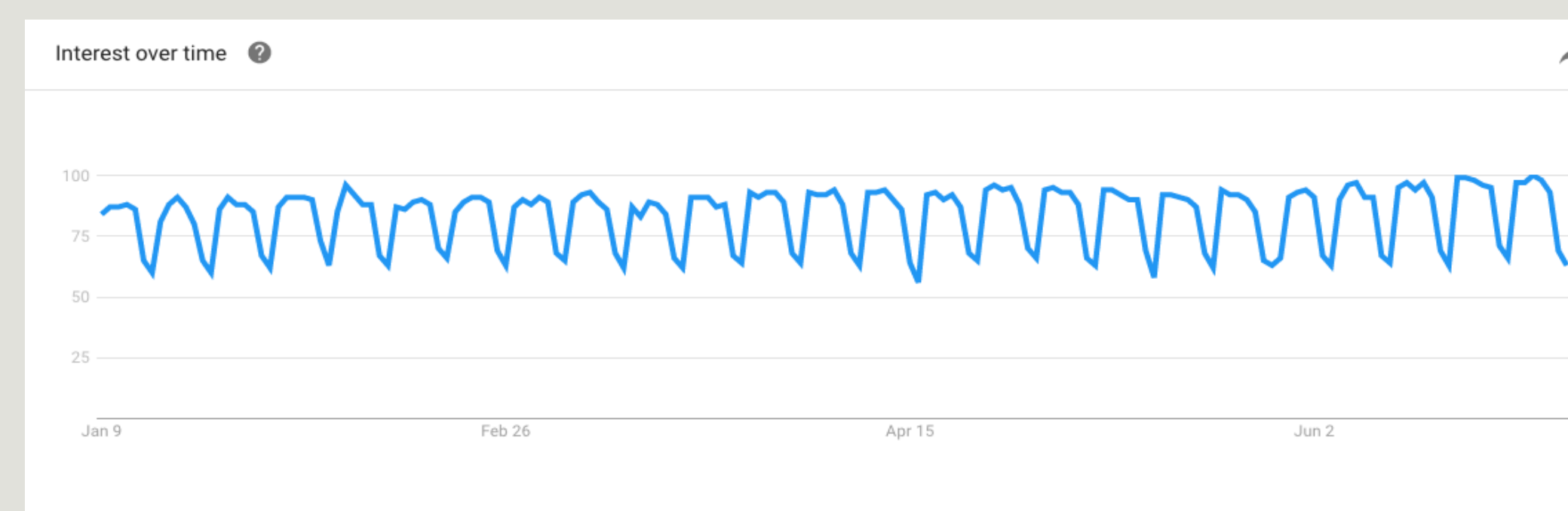


Figure 3. Example of weekend effect on search volume data.

## REFERENCES

Merve Alanyali, Helen Susannah Moat, & Tobias Preis. (2013). Quantifying the Relationship Between Financial News and the Stock Market. *Scientific Reports, 3*(1), 3578.

Tobias Preis, Helen Susannah Moat, & H. Eugene Stanley. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports, 3*(1), 1684.

Tobias Preis, Helen Susannah Moat, H. Eugene Stanley, & Steven R. Bishop. (2012). Quantifying the Advantage of Looking Forward. *Scientific Reports, 2*, 350.

## OUR METHOD

- Refine *Google Trends* strategy to work for different sectors of the market. More targeted sector targeted keywords that included using sector specific data in *Google Trends*.
- Remove weekends our data to account for the systematic effect they have on search volume (Figure 1).
- Institute a threshold which is produced by testing values between 0 and 100 percent and picking that which gives the best return. This thresholds triggers a trade to occur to reduce the effect of small changes in stock price on overall returns.

## CONCLUSION

- Removing weekend data proved effective by causing less days to be seen as increases in search volume. Thus the long position will be taken less often and the short position more often.
- Our research found that applying a threshold number produces significantly better results.
- Keyword and sector specific threshold number outperformed general threshold.

## FUTURE WORK

- Implement more removal of weekends for different sectors of the stock market.
- Incorporate the use of a moving average to mitigate the effect weekend data has on search average.
- Continue application of threshold technique on different sectors of the stock market besides the energy sector.